# Interpretable Network Traffic Classification by Integrating Human Expertise with Machine Learning

Igor Cherepanov

Fraunhofer Institute for Computer Graphics Research IGD

Fraunhofer IGD Darmstadt

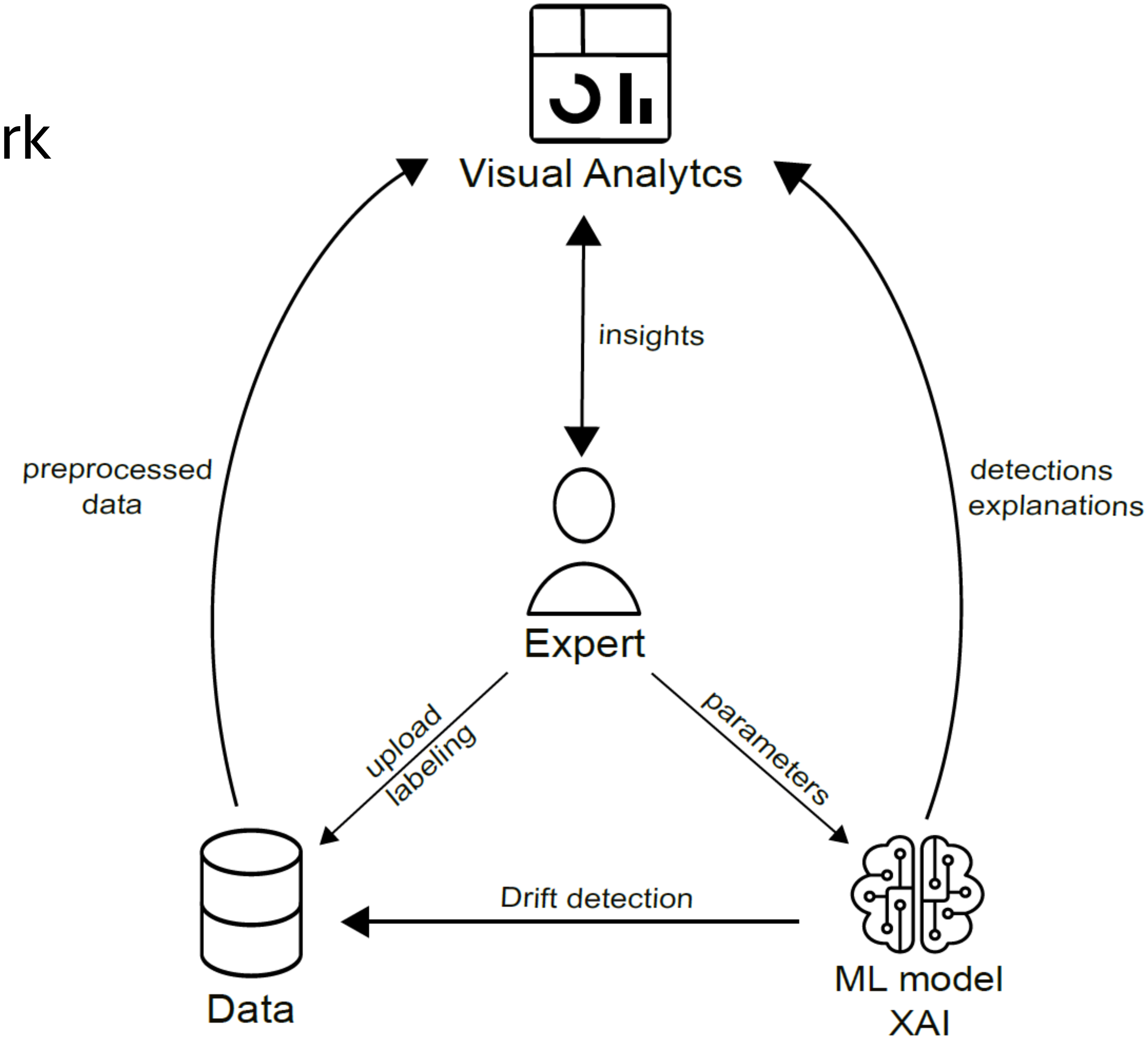
Information Visualization and Visual Analytics (IVA)

Supervised by Prof. Dr.-Ing. Jörn Kohlhammer

# Motivation

- Current situation with ML models
  - Learn hidden relationships in data
  - Excellent performance classification and anomaly detection
- Our goals in fusion with VA:
  - Classification alone is not enough (lacking trust, validation, and understanding)
  - Leverage machine learning for valuable insights
  - Enhance ML systems with domain knowledge

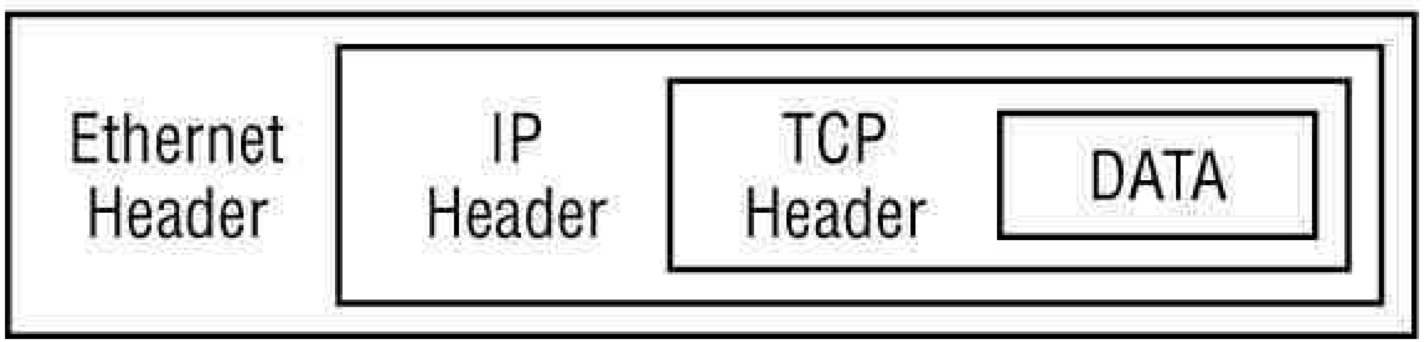
# System framework



# Our case study in cybersecurity

Data

- Network data provided in the form of PCAP files




Network data is often compared to an onion with a header for each layer, followed by the payload.

# Our case study in cybersecurity

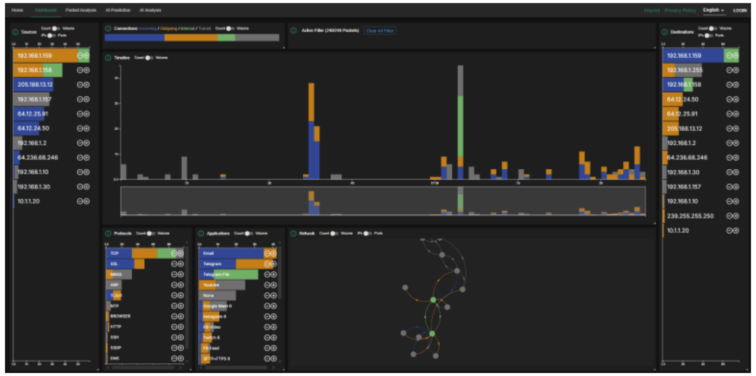
- Data
  - Network data provided in the form of PCAP files
- User
  - Administrators
  - Cybersecurity professionals
  - Malware analysts
- Tasks
  - Develop and enforce security policies and procedures
  - Optimize Quality of Service (QoS)
  - Manage network resources

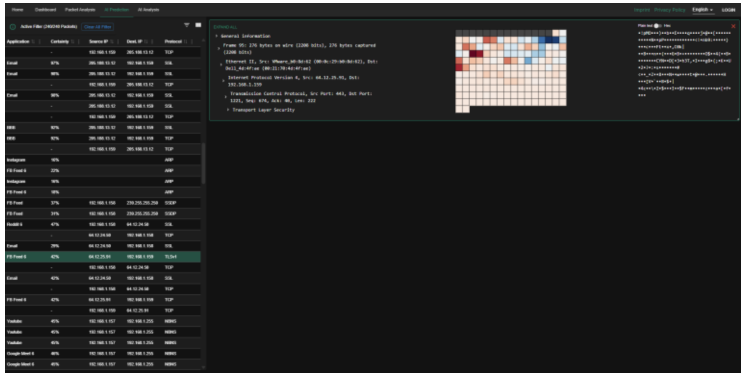
# Progress on our framework

VA

VA+ML+XAI(local)

VA+ML+XAI(global)





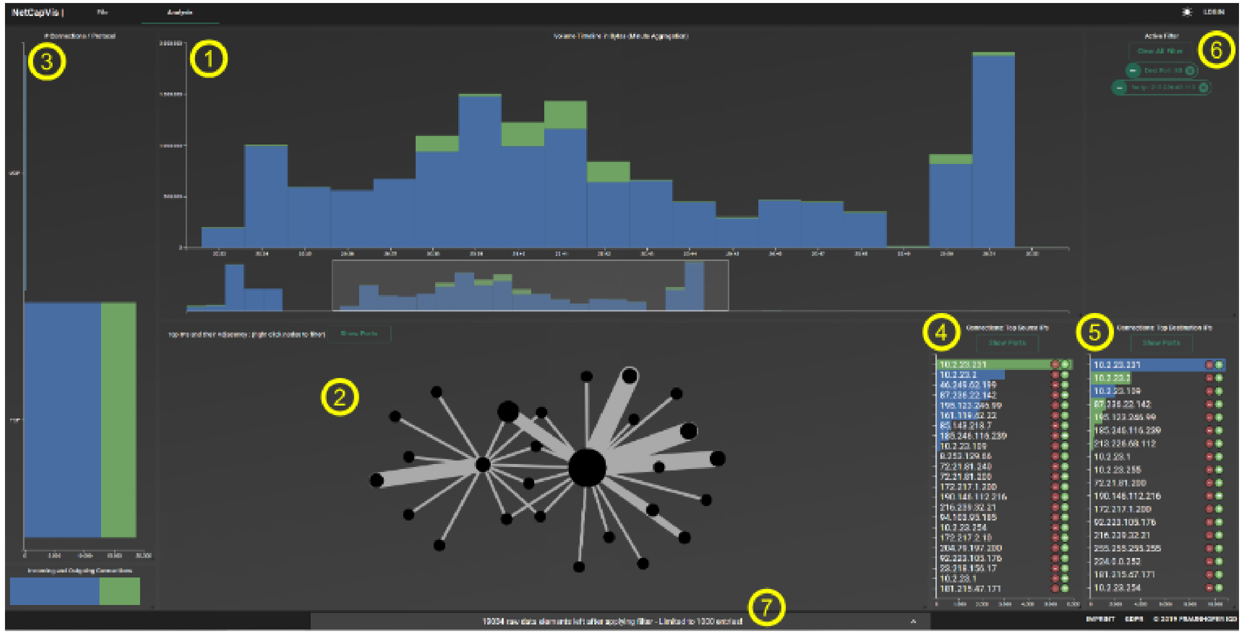
# Visual Analytics for Network Packet Captures

**NetCapVis:**
**Web-based Progressive Visual Analytics for Network Packet Captures**

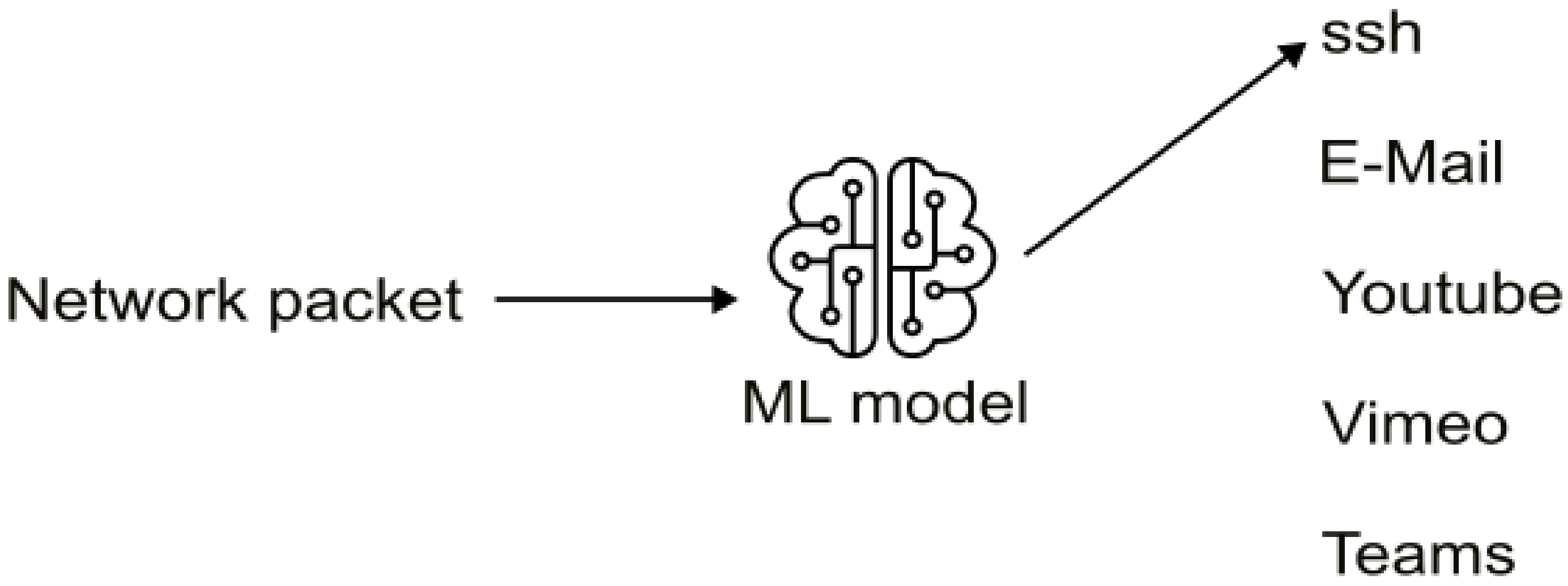
Alex Ulmer* David Sessler† Jörn Kohlhammer‡

Fraunhofer IGD,
Technische Universität Darmstadt, Germany

- Visual analytics system for analyzing PCAP data
- Parsing of PCAP files
- Interlinked visualization of
- network statistics, including:
  - Network graph
  - Timeline view
  - Bar charts
  - Listing of IP addresses
  and protocols



# ML Integration into the Framework

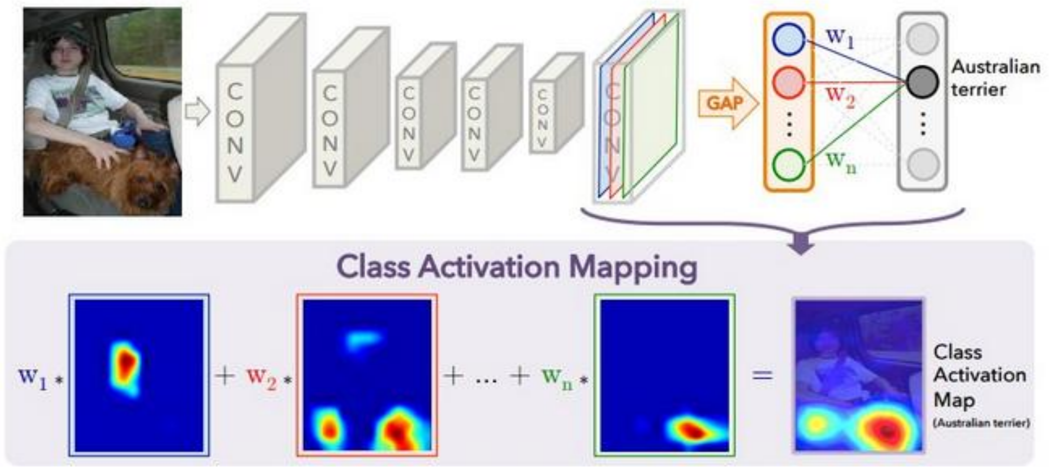


# ML Integration into the Framework

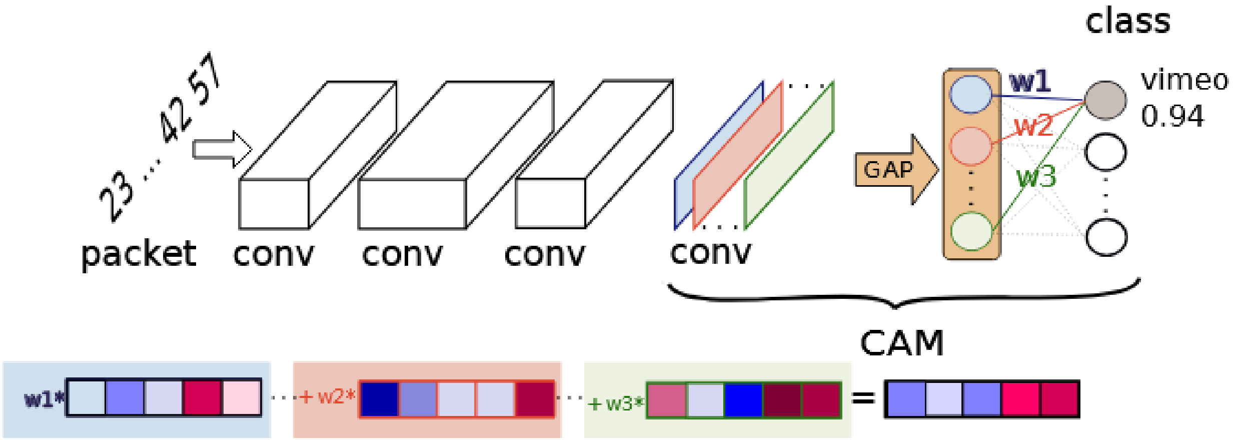
- Researched and compared different ML approaches
  - 1D-CNN model achieved the best performance (confirmed by the field of network research)
- Applied 1D-CNN and evaluated
  - Experimenting with hidden layer depth and width
  - Adjusted convolutional filter sizes
- Created a custom dataset
  - Includes novel and modern application classes (youtube, facebook ...)

# Local XAI Integration into the Framework

- Network packets resemble images, they are one-dimensional, and neighboring bits are frequently related.
- XAI methods from image processing are suitable
  - Class Activation Maps



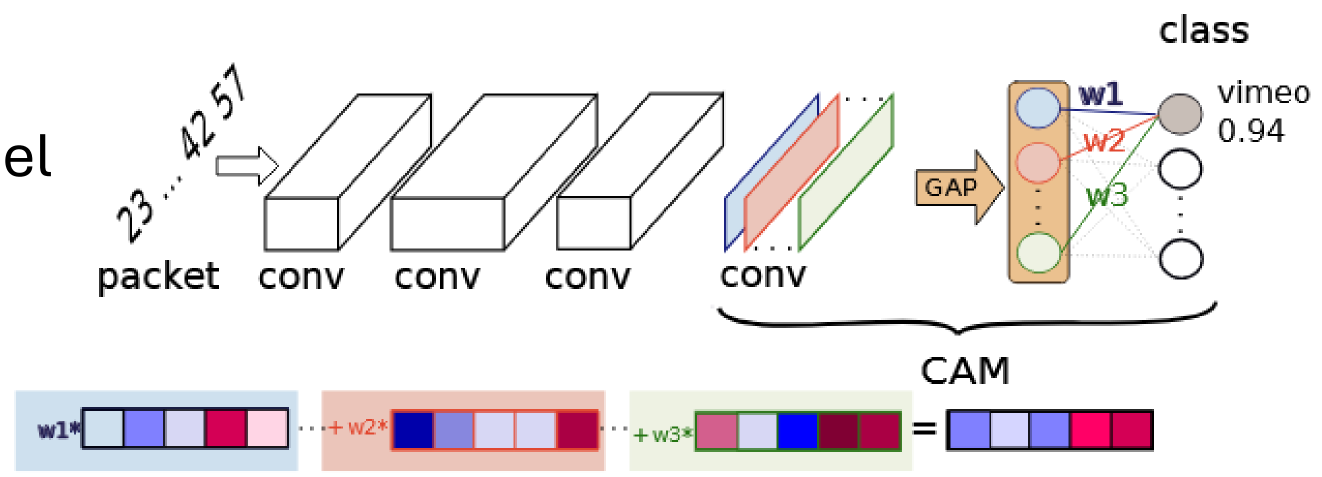

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

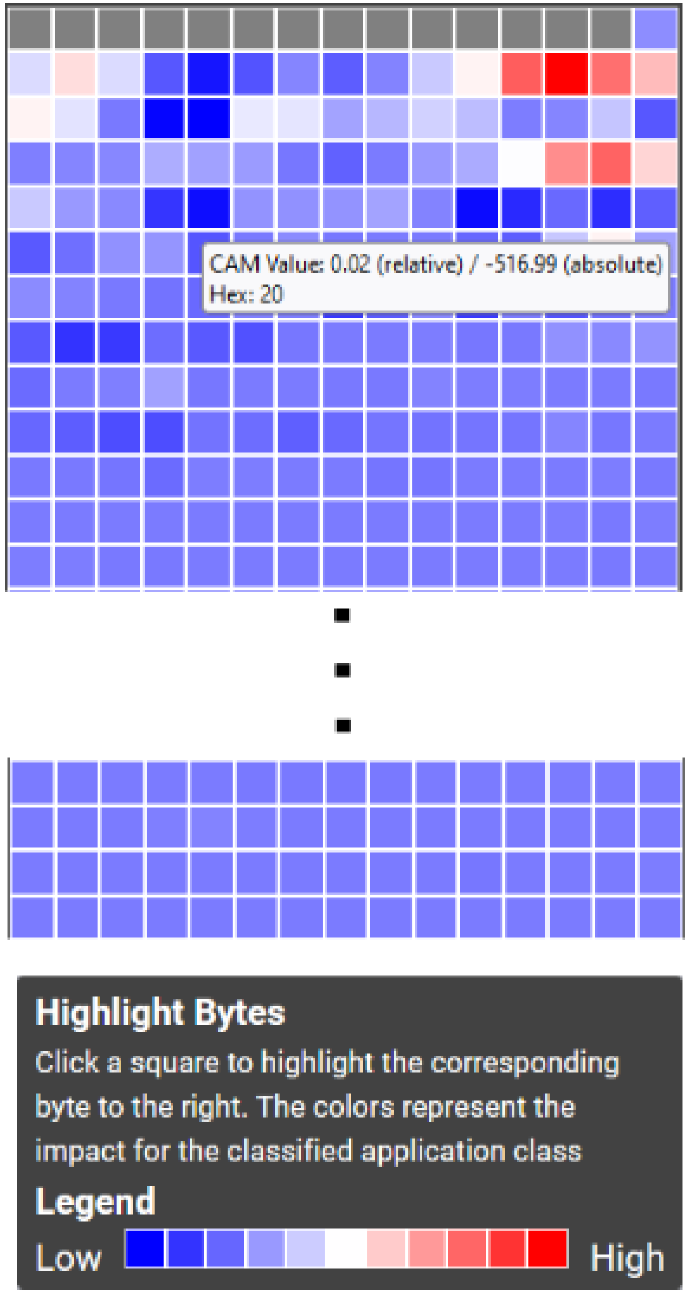


# Local XAI Integration into the Framework

## CAM Calculation:

- The predicted class score is mapped back to the previous convolutional layer to create the CAMs. The resulting CAM is a sum of all convolutional feature maps of the last convolutional layer multiplied by the weights of output layer.
- Computationally cheap
- calculated directly
  on the resulting trained model





CAM Value: 0.02 (relative) / -516.99 (absolute)
Hex: 20
Highlight Bytes
Click a square to highlight the corresponding byte to the right. The colors represent the impact for the classified application class
Legend
Low
High

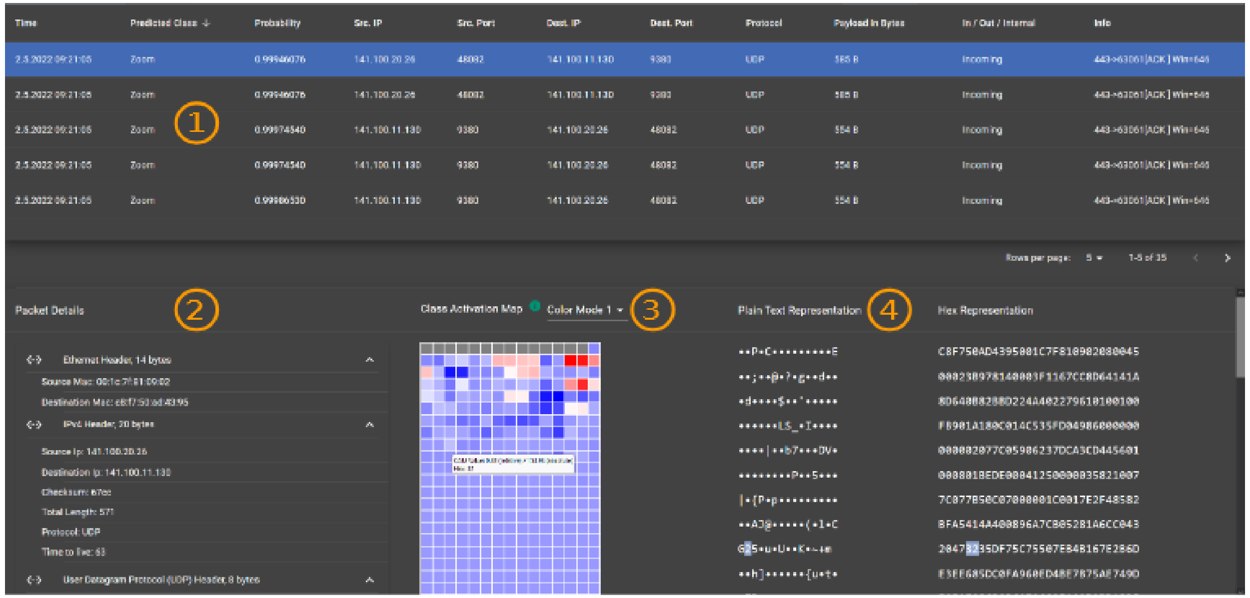
# Local XAI Integration into the Framework

- Overview of all packets with predicted class
- Packets can be viewed in raw format or as a folded tree structure showing all segments
- Investigation of the most impactful bytes in a PCAP file for a predicted application class.

**Visualization Of Class Activation Maps To Explain AI Classification Of Network Packet Captures**

Igor Cherepanov* Alex Ulmer† Jonathan Geraldi Joewono ‡ Jörn Kohlhammer§

Fraunhofer IGD
Technische Universität Darmstadt, Germany

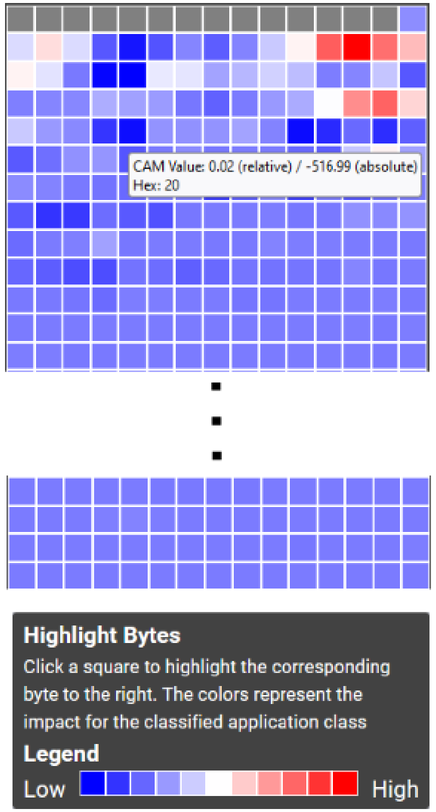


# Local XAI Integration into the Framework

**Findings:**

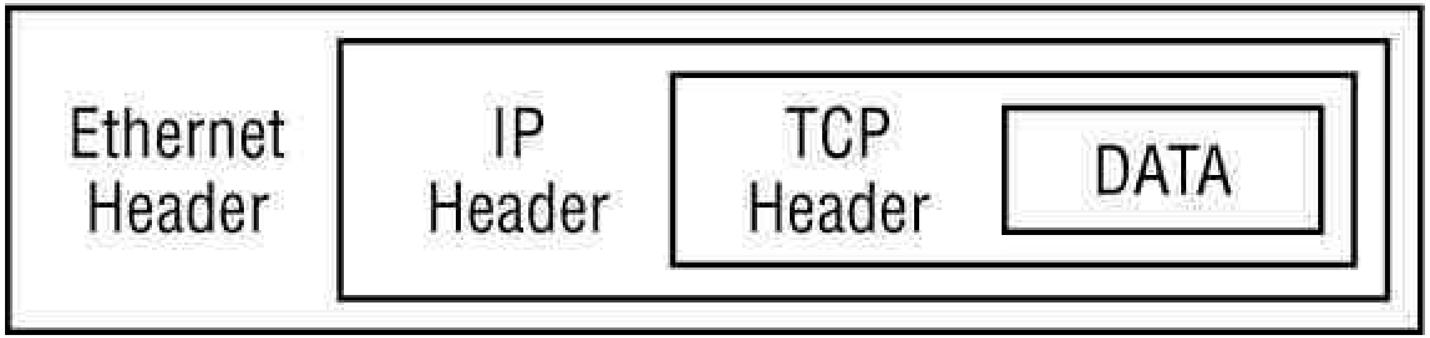
- The model can be simplified without significant performance loss
  - Since the end of the packet is usually encrypted
- Misleading features
  - Sequence number

| Input Size | Accuracy | | |
|---|---|---|---|
| | ISCX2016 | IVA2022 | VNAT2023 |
| 130 | 0.9726 | 0.9872 | 0.9985 |
| 150 | 0.9800 | 0.9853 | 0.9969 |
| 170 | 0.9758 | 0.9887 | 0.9989 |
| 200 | 0.9655 | 0.9836 | 0.9976 |
| 1500 | 0.9586 | 0.9781 | 0.9975 |

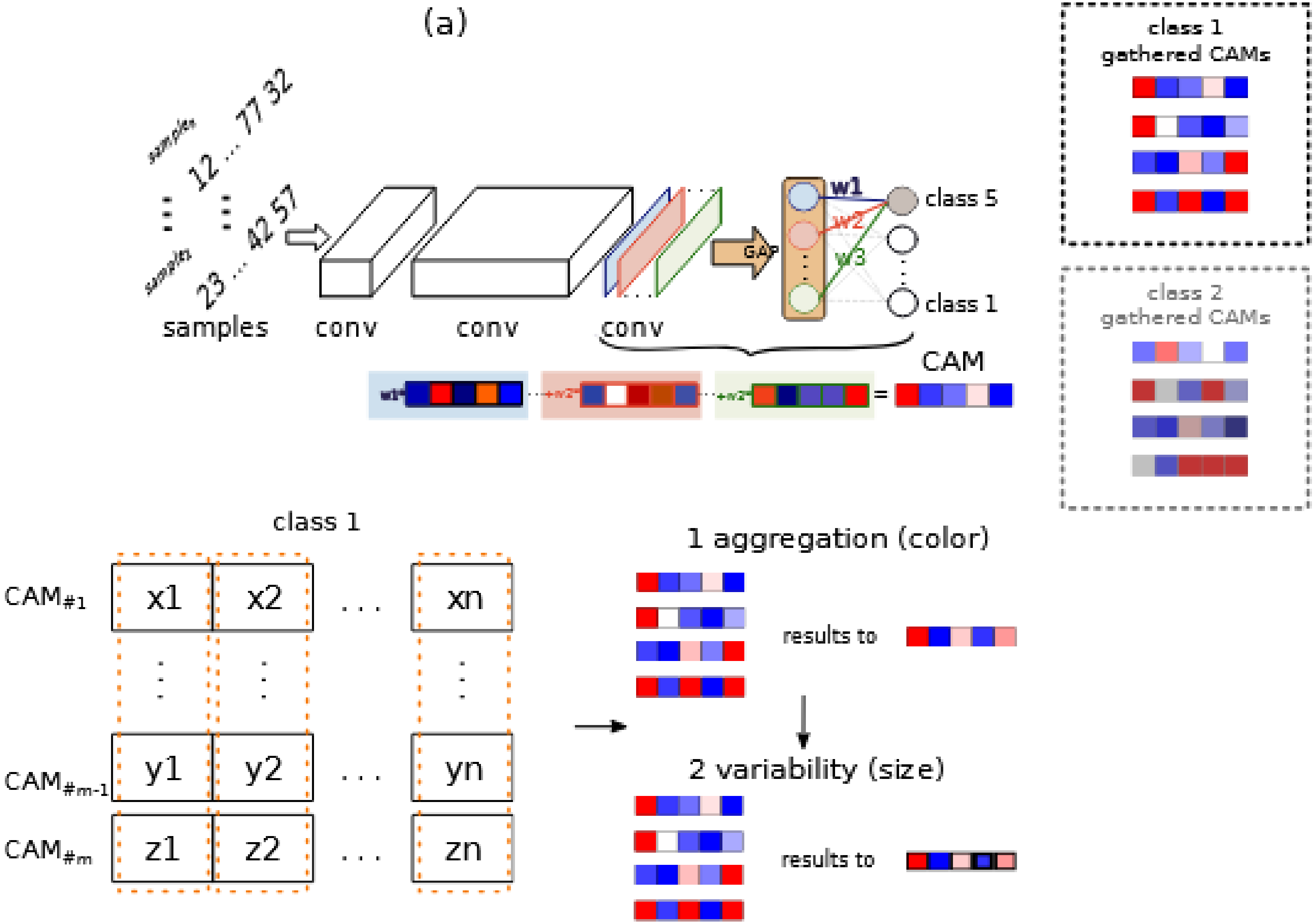


# Global XAI Integration into the Framework

- Position of each unit of information remains constant, lies in its ability to aggregate local explanations for features across all sample



# Global XAI Integration into the Framework

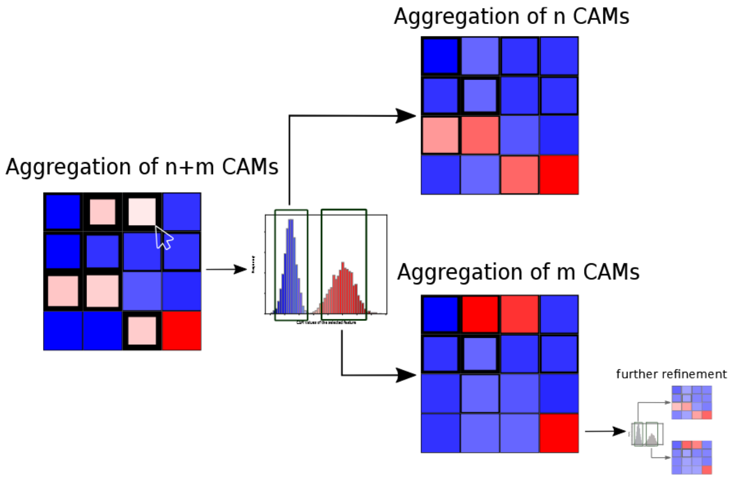


# Global XAI Integration into the Framework

Towards the Visualization of Aggregated Class Activation Maps to Analyse the Global Contribution of Class Features

Igor Cherepanov[1], David Sessler[1], Alex Ulmer[1], Hendrik Lücke-Tieke[1], and Jörn Kohlhammer[1,2]

[1] Fraunhofer IGD, 64283 Darmstadt, Germany
[2] Technische Universität Darmstadt, 64289 Darmstadt, Germany
{igor.cherepanov, david.sessler, alex.ulmer, hendrik.luecke-tieke, joern.kohlhammer}@igd.fraunhofer.de



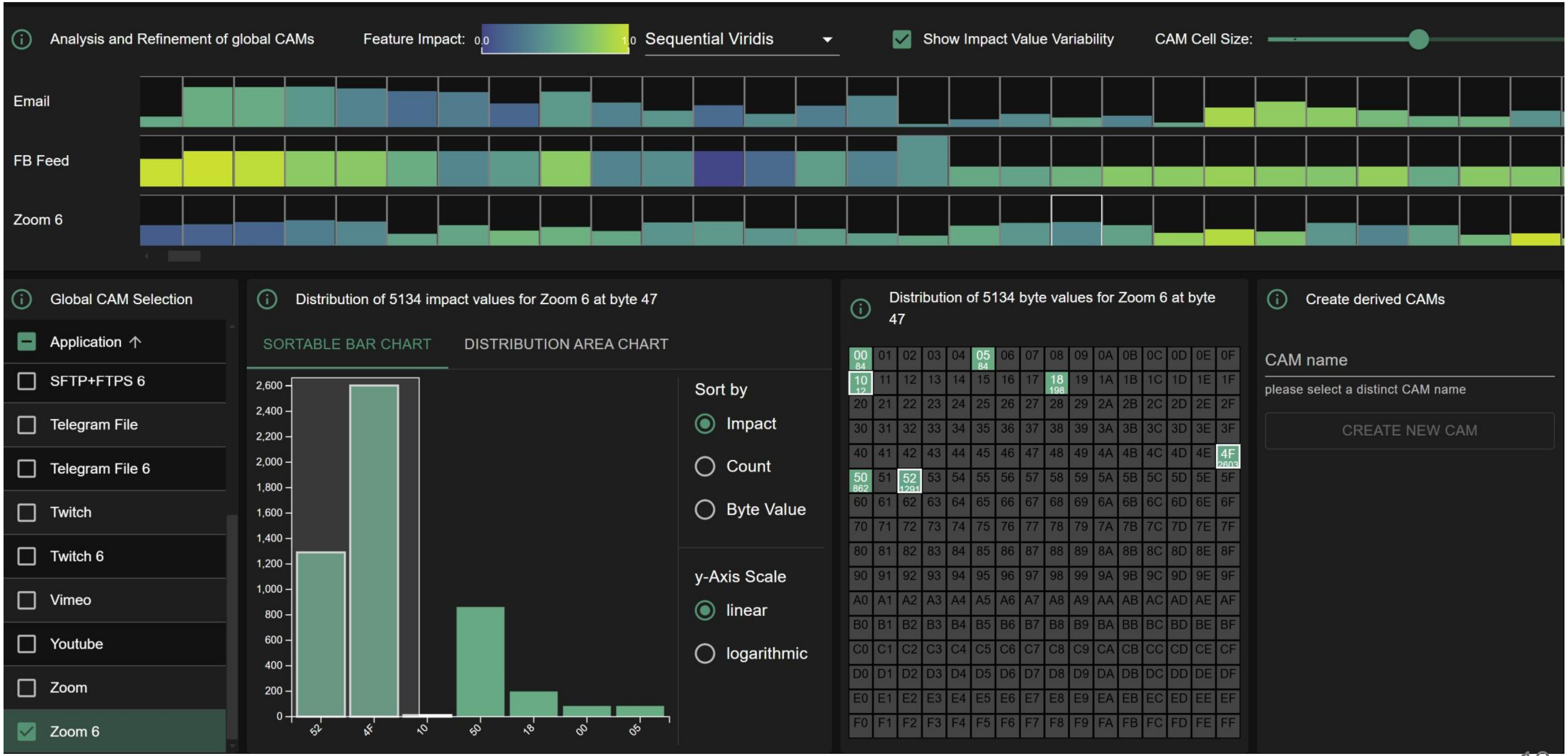

Interactive Analysis of Global Explanations using Aggregated Class Activation Maps for Network Data

I. Cherepanov[1], D. Sessler[1], A. Ulmer [1], F. Wagner[1], T. May [1], J. Kohlhammer[1,2]

[1]Fraunhofer IGD, Germany
[2]TU Darmstadt, Germany


Planned: CGF Submission 2025

# Global XAI Integration into the Framework



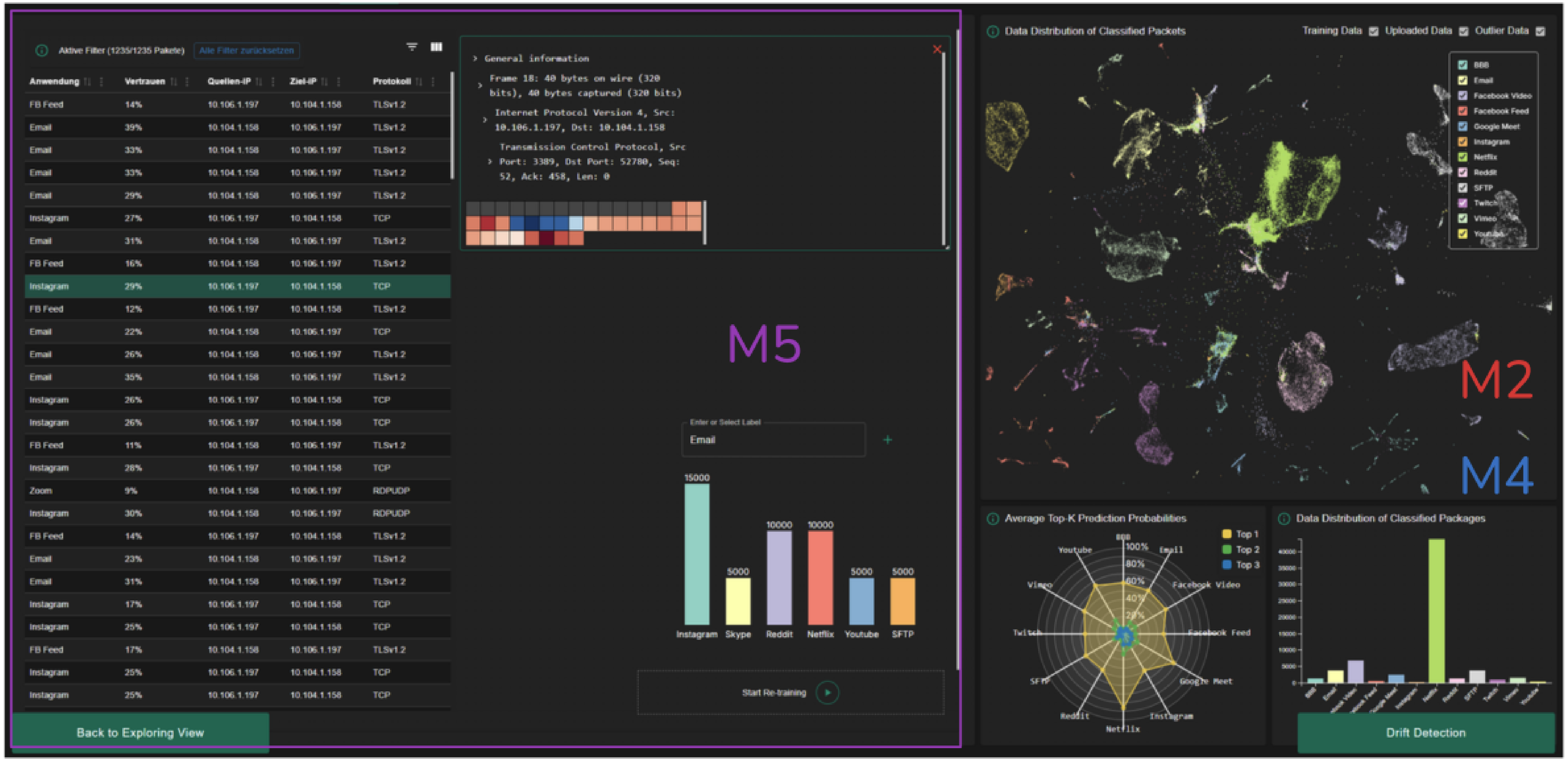
# Global XAI Integration into the Framework

# Enhancement with domain knowledge

Present work:

- Data Drift
  - maintain model accuracy over time
- Evaluation in the domain is challenging due to limited background in simplified ML concepts
  - Challenges in labeling
  - Data drift is difficult for domain experts to understand and interpret
  - Increased complexity when all components (labeling, drift, XAI) are combined

# Enhancement with domain knowledge



# Enhancement with domain knowledge

Future work:

- Model Improvement
  - Applying active learning to iteratively improve the model with informative samples
  - Creation of a more robust model
    - For that we need also a solid dataset as well
- Automatic Rule Extraction/Creation using XAI and raw data
- Providing guidance to help experts overcome the complexity

Thank you for your attention!